Realtime Detection and Tracking of Crowd Sensing for Obtaining Traffic Density

Jazeerah A [1], Fathima Banu H [2], Ansy P K [3], Karthik S [4]
[1, 2, 3] Department of Computer Science and Engineering, University College of Engineering, Nagercoil, Tamil Nadu, India.
[4] Teaching fellow, Department of Computer Science and Engineering, University College of Engineering, Nagercoil, Tamil Nadu, India.

Abstract – Pedestrian safety is a rising social concern nowadays. One important cause affecting pedestrian safety is the existence of temporary obstacles on the sidewalk and using smartphones while crossing in oncoming traffic (e.g. temporary parking, road repairing, cellphone conversation, text messaging).Human crowd motion is mainly driven by self-organized process based on local interactions among pedestrians. In Crowd Watch which leverages mobile crowd sensing and crowd intelligence aggregation to detect temporary obstacles and make effective alerts for distracted walkers. The Dempster–Shafer Evidence Theory is then used to fuse the behavior and visual contexts, and further calculate the confidence of obstacle existence. But the performance of the system would being influenced by some factors such as the participant density, obstacle size, and GPS error. In real-time detection and tracking of crowd sensing using morphological erosion and dilation algorithm aiming at the problems of participant density and low detection accuracy, a fast and accurate detection method based on obtaining the traffic density is proposed using IoT(Internet of things) technique.

Index Terms – object detection, Tracking, Data association, Crowd density, Crowd analysis, Automatic surveillance.

1. INTRODUCTION

The integration of sensing and embedded everyday computing devices at the edge of the Internet will result in the evolution of an embedded Internet or the Internet of Things. The traffic video monitoring and surveillance systems have been widely used in traffic management. Most of the companies have started to use several cameras for the use of traffic surveillance system. The surveillance system extracting useful information such as traffic density, vehicle types from these camera systems due to the high number of cameras in use. Manual analysis of these systems is now inapplicable.

Due to the development of intelligent systems that extract traffic density and vehicle classification information from traffic surveillance systems is complex in traffic management and traffic density. It is important to know that the roads real time especially in megacities for signal control and effective traffic management. Time estimation of reaching from one location to another and recommendation of alternative routes using real time traffic density information are very valuable for mega city residents. Manual analysis of pedestrians and crowds is often impractical for massive datasets of surveillance videos. Automatic tracking of humans is one of the essential abilities for computerized analysis of such videos. In this paper, two prior methods are presented for automatic pedestrian tracking in videos with low and high crowd density. For videos with low density, first each person is detected using a part-based human detector. Then, a global data association method is used which is based on Generalized Graphs for tracking each person in the entire video.

In videos with high crowd-density, the individuals are tracked using a scene structured force model and crowd flow modeling. Also, along with this an another approach is mentioned which utilizes contextual information from the scene without the need to learn the structure of the scene. Performed evaluations show the presented methods outperform the currently available algorithms on several benchmarks. The density of pedestrians significantly impacts their appearance in a video. For instance, in the videos with high density of crowds, people often each other and usually few parts of the body of each individual are visible.

A) Crowd on roadside

B) Crowd on public place

Fig. 1. Traffic density in public places

On the other hand, the full body or a significant portion of the body of each pedestrian is visible in videos with low crowd-density. These characteristics require tracking methods which suite the density of the crowd. In our work, we present two state of the art methods for tracking pedestrians in videos with low and high density of crowds. For videos with low density of pedestrians, first we detect individuals in each video frame using a part-based human detector which efficiently handles occlusion. Later, we employ a global data association method based on Generalized Minimum Clique Graphs for tracking each person over the course of the whole video. We present two approaches to tracking for videos with high density of crowds. In the first one, the scene layout constraint which is captured by learning Dynamic Floor Field, Static Floor Field and Boundary Floor Field along with crowd flow is leveraged to track individuals in the crowd. In the second approach, no learning or crowd flow is used to track targets. Instead, the tracking is performed utilizing the salient and contextual information.

2. RELATED WORK

The main methods used for counting pedestrians in the video stream are the followings. Regression methods. For a video stream, a function of the number of people is built in a particular region depending on the different visual features of the image. Most often as such features the selection of the structural elements, gradient, texture characteristics of the foreground are used(Chan,2008). The foreground is obtained by removing a stationary background and the following correction of perspective distortion. The methods of this class require manual marking and further learning algorithm that introduces significant restrictions. If there are several cameras, training for each camera is not always possible, and the use of learning algorithm for one camera reduces the accuracy on other cameras. Lighting conditions may also vary significantly, and that may affect the accuracy of calculating the number of people. This approach to urban cameras in St. Petersburg has been tested in(Kurilkin,2015). Methods of trajectories clustering. Algorithms of this group consider people as a set of moving points. Points that move close to each other are considered to bean individual. The accuracy of detection of individuals in the crowd is affected by choice of clustering algorithm(Brostow,2006;Rabaud,2006; Sidla,2006; Antonini,2006).Such algorithms face some difficulties incases where people are inactive, waiting inline, are close to the each other in the crowd. Methods of detection of individuals. This group of methods aims to detect individualism the video frame. The algorithm can be used as to detect a human, as for distinctive partsTo count the pedestrians using surveillance cameras, we have chosen a method of detection of individuals. When algorithm applies to detect people in real video stream it should be noted that the detector training for each specific camera, when the number of cameras is large, is not possible. Thus the detectors trained on different data sets, have been tested on real data, which are not included in the training set. For testing, we have selected a few videos recorded with surveillance cameras. For each file, we made a manual calculation of the number of persons crossing the line and then made an automatic calculation using detection algorithm. For all detectors, one tracking algorithm described below is used. Tracking allows determining the location of moving objects in time on video. Tracking algorithm gets the current position of the object identified by the detector, builds the model and performs a search an object in the new frames. The performance of modern detectors and the nature of the data enables to detect the same objects in each frame of a video sequence. To predict and smooth the trajectory of the object a Kalman filter is used. To compare the moving objects in each frame we use a special method based on the minimization of the overlap between the predicted area of bounding box and detected one. Due to permanent changes taking place with an object in motion, such as a change in linear dimensions, illumination, the identification individual characteristics for tracking becomes highly complicated, the current approach allows to rely on the accuracy of the detector, thus minimizing the tracking error, DETER can detect human and track can analyze the trajectory of their motion for threat evaluation.

3. SYSTEM OVERVIEW

Crowd Analysis is the process of understanding the behavioral patterns and density of a particular cluster of a crowd. Using crowd analysis, we can determine the number or type of people at a certain location. We propose an automated system that monitors the crowd and estimates its density. Combining the advantages of Pixel Statistics feature and Texture Analysis, reduces distortion. Generally, an image's regions of interest will focus on objects in its fore-ground. After the stage of image preprocessing object localization is required which may make use of this technique. Background subtraction is a widely used approach for detecting moving objects in videos from static cameras.

The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, often called background image, or background model. Background subtraction is mostly done if the image in question is a part of a video stream. Background subtraction provides important cues for numerous applications in computer vision, for example surveillance tracking or human poses estimation The density of the crowd and their behavioral patterns can be analyzed and understood. This is done by continuous or periodic monitoring of the crowd using cameras.

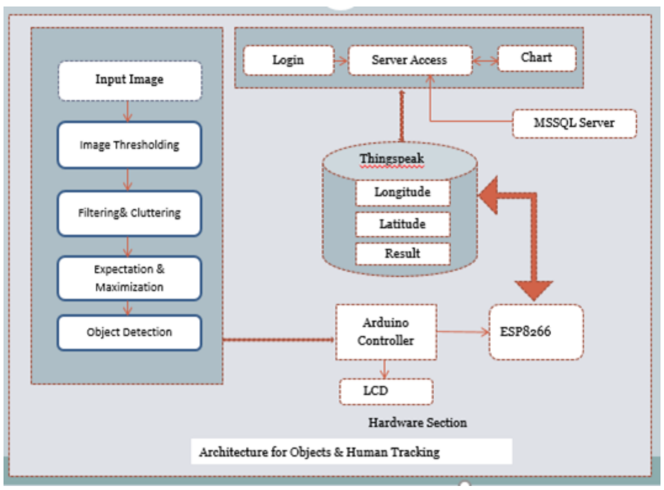

Fig. 2. System Overview of Objects and Human Tracking

1) Video database: In this section, moving object detection was carried out applying the existing Background Subtraction(BS) technique as well as our proposed and implemented Threshold Value(TV) adjusting technique. Thus, selected threshold values minimize noise effects. From various experiments carried out, the optimal threshold value is 25. It can be observed that the image with Threshold Value (TV) 128and 255 are darker than the previous image. Finally, image with TV 255 is completely dark due to increasing the TV up to the peak value. In case of camera pixel noise due to illumination changes, it was removed using opening filter. Small white pixels also have been removed in this process to smoothen the image. Finally motion boundary or blob indicator is used to detect the target objects.

2) Frame Grabbing Module: The frame grabbing module is responsible for dealing with the input device. The input device can be a digital video camera connected to the computer, or a storage device on which a video file or individual video frames are stored. This module abstracts the nature of the input device any from the rest of the system so that changing the input device does not affect the rest of the system.

3) Human Detection Module: This module is responsible for invoking the detection algorithm. Ideally, the detection algorithm is to be run on each input frame. However, this will inhibit the system from meeting its real time requirements. Instead, the detection algorithms in our implementation is invoked every two seconds. The location of the human targets in the remaining time is determined by tracking the detected humans using the tracking algorithm. To further speed up the process, the detection algorithm does not look for humans in the entire frame. Instead, it looks for humans in the regions determined to before ground regions. To determine the foreground regions, a stabilization algorithm is used to align the current frame with a preceding frame and with a succeeding frame. The current frames are subtracted from the two other frames after the process of alignment. The result of each subtraction is thresholder to form a binary image that represents the locations of foreground objects in the two subtracted frames. The locations of the foreground objects in the current frame can be known by using the results of the two subtractions which are combined by an AND operation. The subtraction is performed in the hue channel of the HSV color space.

4) Object Tracking Module: In this module, the frames and detections received from the human detection module are processed, and the information are retrived regarding all the existing tracks. Whenever a new frame is received, making use of the new bounding boxes, the already existing tracks are extended by locating their locations for each track in the frame. If received frames are present with new detections, the new detections are compared to the already existing tracks for retrieving the differences and the significant overlaps of the new detections with any of the existing tracks are ignored. Else, a new track is created for this new detection. A track is discontinued if the tracking algorithm fails to extend it in a newly coming frame.

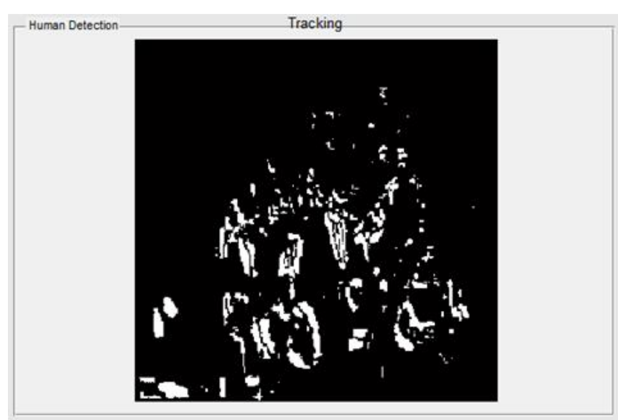

Fig. 3: Human and object tracking.

5) Motion Analysis Module: When the length of a track exceeds some specific length, typically, two seconds, the motion analysis module is invoked. The motion analysis

module analyzes the periodicity encountered in the track. Based on the result of this analysis, it decides whether the tracked object is indeed a human or not. Thus, using this method the detection results can be double checked by the motion analysis. In our experiments, that results in a reduction in the false positives produced by the detector.

6) Output Module: When the motion analysis module completes the analysis of each track element, or removes the track element because of being too short to be analyzed, this frame is passed to the output module. The output module marks the detected human locations in the frame and sends it to the output device, which can be the display monitor or a storage device.

7) Human Detection and Object Tracking Modules: This interface is a queue of structure elements. Each element contains a pointer to the frame along with a list of detections found in it. If the list of detections is empty, then, either the detection algorithm was not run on this frame, and detection algorithm did not find any human in it.

8) Object Tracking and Motion Analysis Modules: Two lists are maintained in the shared data structure which are a frames queue, which is a queue of frames queue elements, and a tracks list which is a list of pointers to track element lists. Each of the frames queue element comprises a pointer to a frame and also a counter that holds the number of objects that are in that frame. In the track list, each track element is a list of track elements. A track element is used to represent the location of the tracked object in one of the frames, which is pointed to by an element in the frames queue. A track element contains two items: a bounding box that specifies the location of this track's object in the corresponding frame, and a pointer to the entry of this frame in the frames queue. When the motion analysis module processes a track, it removes all its track list elements except for the most recent one; so that it can be tracked in the next frame.

9) Motion Analysis and Output Modules: The motion analysis module and the output module are interfaced with the help of simply a queue of pointers to frames that has become ready for output. The Motion Analyzer Module is responsible for sending frames ready for output to the output module, along with the bounding boxes that identify the targets that have been verified to be humans.

Algorithm Expectation and maximization

{Input s (discriminant scores from

Identified spectra), €}

{Let p be the N-vector of the probabilities

For each identified spectra being a member

of the incorrect distribution}

i ←1

$\hat{\pi}0,i,\hat{\mu}i,\hat{\sigma}i,\hat{\alpha}i,\hat{\beta}i,\gamma I$ ←initialize($\hat{\pi}0,,\hat{\mu},\hat{\sigma},\hat{\alpha},\hat{\beta},\gamma$)

convergence ← FALSE

WHILE convergence == FALSE **DO**

i ← i + 1

{E-Step}

p←EstimateLikelihoodIncorrectMembership($\hat{\pi}0,i,\hat{\mu}i,\hat{\sigma}i,\hat{\alpha}i,\hat{\beta}i,\gamma i,s$)

{M-Step : Update the parameters based on p}

$\hat{\pi}0,i,\hat{\mu}i,\hat{\sigma}i,\hat{\alpha}i,\hat{\beta}i,\gamma i$

←EstimateParametersFromLikelihoods(p,s)

IF | $\hat{\pi}0,i,\hat{\mu}i,\hat{\sigma}i,\hat{\alpha}i,\hat{\beta}i,\gamma I$ - $\hat{\pi}0,i\text{-}1,\hat{\mu}i\text{-}1,\hat{\sigma}i\text{-}1,\hat{\alpha}i\text{-}1,\hat{\beta}i\text{-}1,\gamma i\text{-}1$|<€ **THEN**Convergence ←TRUE

END IF

END WHILE

p←EstimateLikelihoodIncorrectMembership($\hat{\pi}0,i,\hat{\mu}i,\hat{\sigma}i,\hat{\alpha}i,\hat{\beta}i,\gamma i,s$)

RETURN $\hat{\pi}0,i,\hat{\mu}i,\hat{\sigma}i,\hat{\alpha}i,\hat{\beta}i,\gamma i,p$

After initialization, the algorithm makes time-lapse video processing. Each frame passes image enhancement procedures with pedestrian detection. Further process each frame processing can be divided into several stages:

A) The prediction of the new positions of the existing tracks. After detecting pedestrians in the frame of video, the algorithm carries outa prediction of the new provisions of the existing tracks. The Kalman filter predicts the centroid of the object of each track in the current frame concerning the previous one and update position. It is assumed that the size of the bounding box in the current frame doesn't change compared to the previous frame.

B) Assignment of objects to tracks. Assigning tracks to detected objects is carried out by minimizing the cost. The value is calculated as a function of the overlap of the predicted and detected box. Price has a minimum value when the predicted box fits the detection. For improving the accuracy, algorithm ignores unrealistic matching.

C) Each assigned track is updated with the last detected object. The algorithm adjusts the approaching of the position of the object with the new detection. Further, the averaging of sizes of the last four frames of the track allows stabilizing the bounding box.

D) Updating the unassigned tracks. Each unassigned track is marked as invisible. The algorithm updates the parameters such as the age of track frames, the number of frames, during which the object detection is not performed.

E) Removal of the lost tracks. Removal of lost tracks is made by a set of rules. The following parameters are taken into account: age, number of frames, during which the object is visible and invisible, the degree of reliability of track.

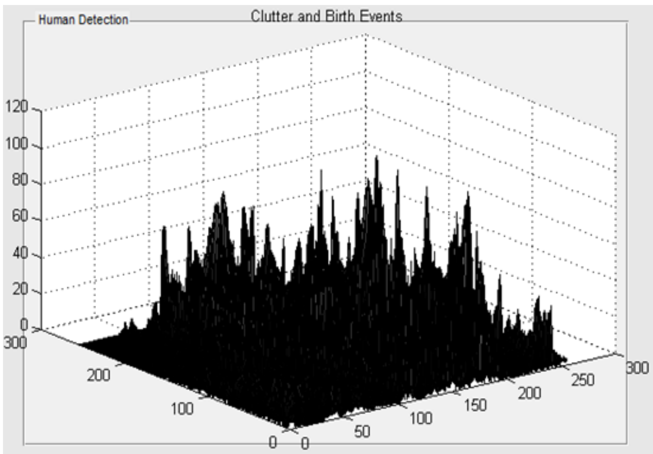

Fig. 4. Histogram of Human Detection

F) The creation of new tracks. New tracks are created from unassigned detection. Each unassigned detection considered as the beginning of a new track. The new track will not be visible while the condition of displaying is false.

G) Processing of the results. For each frame, the algorithm may show bounding boxes of detected pedestrians, obtained tracks, as well as additional information. Also at this stage, counting of A comparison of methods to detect people flow using video processing Alexey V. Kurilkin and Sergey V. Ivanov 128 the number of pedestrians which have passed given the line, may be done. It is thought that the line is passed if the track crosses a line in one direction. As a result, the algorithm performs the detection of individual tracks of pedestrians and count a total number of people passing through the line in a certain direction per time unit.

Fig 5. Crowd Count

4. PERFORMANCE EVALUATION

To evaluate the effectiveness of the proposed object identification system, we divide videos into three types based on the level of movements. These three types are videos with frequent movements videos with intermediate movements and videos with infrequent movements. We conduct a comparative experiment as an example incorporating hint information or without incorporating hint information.

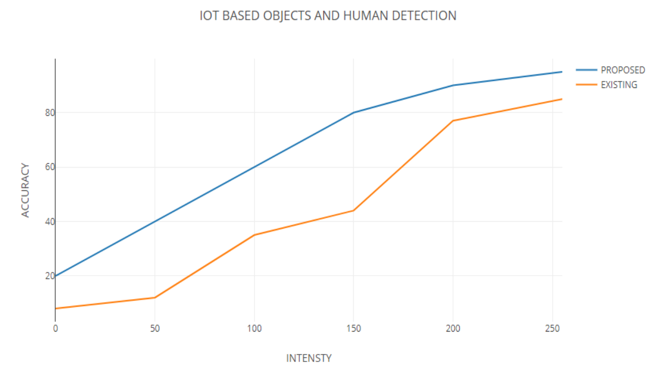

Fig. 6. Performance Evaluation on Accuracy

For an object detection and tracking task the errors that can affect the metric scores can be due to a single or a combination of the following errors - spatial inaccuracy, temporal inaccuracy and missed detects. To measure the influence of all of these factors at the same time will not reflect the behavior of the measures to individual errors. There are many aspects of an algorithm that affect the final scores of the detection and the tracking measure.

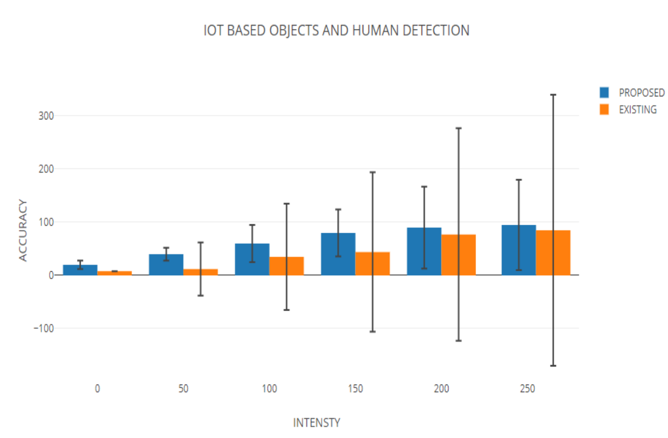

Fig.7: Error chart

5. CONCLUSION

In this paper, we propose an intelligent system for tracking a human and objects in the surveillance camera and pass the values through IOT. Which can be applied in home and business surveillance system consisting of static camera. It should find the traffic in particular area. Proposed method used to separate the crowd of the person count from vehicle among the moving object. It is possible to monitor wide range of area with minimum number cameras and track a particular moving object among many ones. The result indicate that our methods to temporary object detection are effective.

REFERENCES

[1] D. Gavrila. Pedestrian detection from a moving vehicle. In ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part II, pages 37–49, London, UK, 2000. Springer-Verlag.

[2] A. Hampapur, L. Brown, J. Connell, , A. Ekin, N. Hass, M. Lu, H. Merkl, S. Pankanti, A. Senior, C.-F. Shu, and Y. L. Tian. Smart video surveillance. IEEE Signal Processing Magazine, pages 38–51, March 2005.

[3] A. Hampapur, L. Brown, J. Connell, N. Hass, M. Lu, H. Merkl, S. Pankanti, A. Senior, C.-F. Shu, and Y. Tian. S3-r1: The ibm smart surveillance system-release 1. In ACM SIGMM workshop on Effective telepresence, 2004.

[4] I. Haritaoglu, D. Harwood, and L. Davis. w 4 : Realtime surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8):809–830, 2000.

[5] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, 34:334–352, 2004.

[6] T. B. Moeslund and E. Granum. A survey of computer vision-based human motion capture. Computer Vision and Image Understanding, 81:231–268, 2001.

[7] V. Morellas, I. Pavlidis, and P. Tsiamyrtzis. Deter: Detection of events for threat evaluation and recognition. Machine Vision and Applications, 15:29–45, 2003.

[8] Honda. Honda Demonstrates Advanced Vehicle-to-Pedestrian and Vehicle-to-Motorcycle Safety Technologies. http://www.honda.com/newsandviews/article.aspx?id=7352-en,August 2013.

[9] K. Dhondge, S. Song, Y. Jang, H. Park, S. Shin, and B. Y. Choi. "Video: Wifi-honk: Smartphone-based beacon stuffed wificar2x-communication system for vulnerable road user safety." In Proceedings of ACM MobiSys'14, 2014, pp. 387–387.

[10] A. Fackelmeier, C. Morhart, and E. Biebl. "Dual frequency methods for identifying hidden targets in road traffic," In Proceedings of Advanced Microsystems for Automotive Applications, 2008, pp.11-20.

[11] L. Shangguan, Z. Yang, Z. Zhou, X. Zheng, C. Wu , and Y. Liu. "CrossNavi: Enabling real-time crossroad navigation for the blind with commodity phones." In Proceedings of UbiComp'14, 2014, pp.787-798.

[12] T Wang, G Cardone, A Corradi, A Corradi, L Torresani, and A. T. Campbell, "WalkSafe: a pedestrian safety app for mobile phone users who walk and talk while crossing roads." In Proceedings of the Thirteenth International Workshop on Mobile Computing Systems and Applications, 2012.

[13] S. Jain, C. Borgiattino, Y. Ren, M. Gruteser, and Y. Chen, "On the limits of positioning-based pedestrian risk awareness." In Proceedings of MARS'14, 2014, pp. 23–28.

[14] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Yen, R. Huang, and X. Zhou. "Mobile crowd sensing and computing: The review of an emerging Human-Powered sensing paradigm." ACM Computing Surveys, 48(1), pp.1-31, 2015.

[15] T. Gandhi and M.M. Trivedi. "Pedestrian protection systems: Issues, survey, and challenges." IEEE Transactions on Intelligent Transportation Systems, vol. 8, no. 3, pp. 413-430, 2007.

[16] B. Guo, Z. Wang, Z. Yu, Y. Wang, N.Yen, R. Huang, and X. Zhou. "Mobile crowdsensing and computing: The review of an emerging human powered sensing paradigm". ACM Computing surveys, 48(1).pp.1-31.2015